

AD-A145 402	AN ITERATIVE PROCEDURE FOR OBTAINING I-PROJECTIONS ONTO THE INTERSECTION O. (U) IOWA UNIV IOWA CITY DEPT OF STATISTICS AND ACTUARIAL SCIENCE. R L DYKSTRA JUN 84	1/1
UNCLASSIFIED	TR-106 N00014-83-K-0249	F/G 12/1 NL

AN ITERATIVE PROCEDURE FOR OBTAINING I-PROJECTIONS ONTO
THE INTERSECTION OF (U) IOWA UNIV IOWA CITY DEPT OF
STATISTICS AND ACTUARIAL SCIENCE.. R L DYKSTRA JUN 84

1/1

UNCLASSIFIED

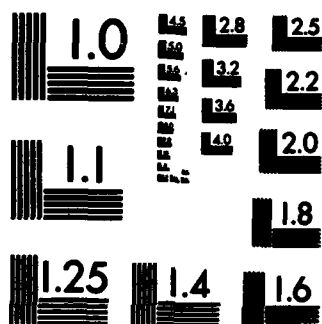
TR-106 N00014-83-K-0249

F/G 12/1

NL

12

END



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A145 402

12

AN ITERATIVE PROCEDURE
FOR OBTAINING I-PROJECTIONS
ONTO THE INTERSECTION OF CONVEX SETS

Richard L. Dykstra

Department of Statistics and Actuarial Science
The University of Iowa
Iowa City, Iowa 52242

Technical Report #106
June 1984

DTIC
ELECTE
SEP 12 1984
S B

DTIC FILE COPY

This work was partially supported by ONR Contract N00014-83-K-0249.

DISTRIBUTION STATEMENT A

Approved for public release
Distribution Unlimited

84 09 06 054

ABSTRACT

A frequently occurring problem is to find that probability distribution (PD) lying within a ^{convex} set δ which minimizes the I-divergence between it and a given PD R. This is referred to as the I-projection of R onto δ . Csizar (1975) has shown that when $\delta = \bigcap_{i=1}^t \delta_i$ is a finite intersection of closed, linear sets, a cyclic, iterative procedure which projects onto the individual δ_i must converge to the desired I-projection on δ providing the sample space is finite.

This document
~~Here we propose~~ an iterative procedure which requires only that the ~~Set~~ be convex (and not necessarily linear) which under general conditions will converge to the desired I-projection of R onto $\bigcap_{i=1}^t \delta_i$.

Key words and phrases: I-divergence, I-projections, convexity, Kullback-Liebler information number, cross entropy, iterative projections, iterative proportional fitting procedure.

AMS 1970 Subject classification: Primary - 90C99

Secondary - 49D99

Short Title: I-Projections on Convex Sets.

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By <i>PER CALL JC</i>	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
<i>A-1</i>	



AN ITERATIVE PROCEDURE
FOR OBTAINING I-PROJECTIONS
ONTO THE INTERSECTION OF CONVEX SETS

1. Introduction

Suppose P and Q are probability measures defined on subsets of the finite set \mathcal{X} , which WLOG we take to be the first m positive integers. The I -divergence of P with respect to Q , also called the Kullback-Leibler information number, cross entropy between P and Q , information for discrimination, entropy of P relative to Q , etc., is given by

$$(1.1) \quad I(P\|Q) = \sum_{k=1}^m p(k) \ln \frac{p(k)}{q(k)}$$

where $p = (p(1), \dots, p(m))$ and $q = (q(1), \dots, q(m))$ denote the probability mass functions (PMF) of P and Q respectively. (In general, we shall indicate a measure on subsets of \mathcal{X} by an upper case letter (P, Q, R, S) and the associated Radon-Nikodym derivative with respect to counting measure (PMF) by the corresponding lower case letter (p, q, r, s).)

We mention that $I(P\|Q)$ is defined analogously for general probability measures on infinite spaces, but for simplicity, we will only consider finite sets. (See Kullback (1959), or Csiszar (1975) for the general definition.)

It is well known that $I(P\|Q) \geq 0$, and that $I(P\|Q) = 0$ iff $P \equiv Q$. Thus it is heuristically reasonable to think of $I(P\|Q)$ as representing a "distance" between P and Q . However, $I(\cdot\|\cdot)$ is not a metric, nor

is the symmetrized version

$$J(P\|Q) = \frac{I(P\|Q) + I(Q\|P)}{2}$$

used by Jeffreys (1948).

Nevertheless,

$$(1.2) \quad \sum_{k=1}^m |p(k) - q(k)| \leq [2 I(P\|Q)]^{1/2},$$

as shown independently by Kullback (1967), Kemperman (1967), and Csiszar (1967), so that we have some idea of what small values of $I(P\|Q)$ imply.

If we interpret $I(P\|Q)$ as distance, it seems natural to define the I -projection of the probability distribution (PD) R onto a set \mathcal{G} of probability distributions as being a PD $Q \in \mathcal{G}$ such that $I(Q\|R) < \infty$ and

$$(1.3) \quad I(Q\|R) = \min_{P \in \mathcal{G}} I(P\|R).$$

In some sense, Q is the PD closest to R that lies within \mathcal{G} .

Minimization problems of the form (1.3) play a key role in the information-theoretic approach to statistics (e.g. Kullback (1959), Good (1963), etc.) and also occur in other areas such as the theory of large deviations (Sanov (1957)) and maximization of entropy (Rao (1965) and Jaynes (1957)).

However, in statistical circles, I -projections are probably most important for being dual problems to certain log-linear model maximum likelihood estimation (MLE) problems. In particular, it is known that the multinomial MLE problem

$$(1.4) \quad \begin{array}{l} \text{Max} \\ p(k) \geq 0, \forall k \\ \sum_{k=1}^m p(k) = 1 \\ \ln p \in \mathcal{M} \end{array} \prod_{k=1}^m p(k)^{n(k)}$$

(where \mathcal{M} is some subspace of \mathbb{R}^m containing the constant vectors) has precisely the same solution as the I-projection problem

$$(1.5) \quad \begin{array}{l} \text{Min} \\ P: S-P \in \mathcal{M}^\perp \end{array} I(P \| U)$$

(where $s(k) = n(k) (\sum_{i=1}^m n(i))^{-1}$, and $U(k) = \frac{1}{m}$ is the uniform PMF). Note that for a subspace spanned by the vectors a_1, \dots, a_t , we have

$$\mathcal{M}^\perp = \{a_1, \dots, a_t\}^\perp = \bigcap_{i=1}^t \{a_i\}^\perp,$$

and hence (1.4) is equivalent to

$$\inf_{P \in S - \mathcal{M}^\perp} I(P \| U) = \inf_{P \in \bigcap_{i=1}^t (S - \{a_i\}^\perp)} I(P \| U).$$

Since the $S - \{a_i\}^\perp$ are linear spaces, Csiszar's algorithm of cyclic, iterated I-projections is appropriate here. In this format, it is easy to see the connection between Csiszar's procedure and the IPFP (iterative proportional fitting procedure) which has received so much attention in the general area of categorical data. Meyer (1980) has an extensive discussion and several examples where he relates Csiszar's procedure and the general IPFP.

Suppose now that $\mathcal{M} = K_1 + \dots + K_t$ is a closed, convex cone expressible as a direct sum of closed, convex cones containing the constant vectors rather than a

direct sum of subspaces. Such a configuration would arise naturally if one were considering order constraints in a log-linear model. Then it is well known that the dual (polar) cone of \mathcal{M} , defined as

$$\mathcal{M}^* = \{y; \sum_{i=1}^m y(i)x(i) \leq 0 \text{ for all } x \in \mathcal{M}\}$$

is expressible as

$$\mathcal{M}^* = (K_1 + K_2 + \dots + K_t)^* = K_1^* \cap K_2^* \cap \dots \cap K_t^*.$$

In unpublished results, Dykstra has generalized the (1.4)-(1.5) duality results to the case where \mathcal{M} is a closed, convex cone (with \mathcal{M}^* replacing \mathcal{M}^\perp in (1.5)). This means that many MLE problems involving partial orders in log-linear models are equivalent to I-projection problems of the form

$$\begin{array}{ll} \text{Min} & I(P||U). \\ P \in & \bigcap_1^k (S - K_1^*) \end{array}$$

If the K_1^* 's are not subspaces, Csizsar's cyclic, iterated scheme need not work. However, the procedure described in this paper will work since the sets $S - K_1^*$ will be closed, convex sets of PD's.

Of course we would really like to be able to identify structure in these log-linear model situations, which leads to the area of inference for various competing models. While these are important questions, we shall only be concerned with the MLE problem in this paper.

Csizar (1975) discusses I-projections in great detail, and has a "geometric" development for I-projections which is quite appealing. (Cencov (1972) also has a geometric development of I-projections, but with the arguments interchanged.) Csizar also discusses the exist-

ence of I projections, and shows that if \mathcal{G} is a convex set of PD's which is variation closed, the unique existence of the I -projection of R onto \mathcal{G} is guaranteed provided there exists a $P \in \mathcal{G}$ such that $I(P\|R) < \infty$. This result is clearly applicable for closed, convex sets of PD's on the finite set \mathcal{X} . We shall make repeated use of the elegant characterization of I -projections given in the following theorem.

Theorem 1.1 (Csiszar). A PD $Q \in \mathcal{G}$ (such that $I(Q\|R) < \infty$) is the I -projection of R onto the convex set \mathcal{G} of PD's iff

$$(1.4) \quad I(P\|R) \geq I(P\|Q) + I(Q\|R) \quad \forall P \in \mathcal{G}.$$

Note that in our setting of finite \mathcal{X} , it follows from (1.1) that (1.4) is equivalent to

$$(1.5) \quad \sum_{k=1}^m (p(k) - q(k)) \ln \left(\frac{q(k)}{r(k)} \right) \geq 0 \quad \forall P \in \mathcal{G}.$$

If in fact Q is an algebraic inner point of \mathcal{G} , i.e. for every $P \in \mathcal{G} (P \neq Q)$, there exists $0 < \alpha < 1$ and $P' \in \mathcal{G}$ such that $Q = \alpha P + (1-\alpha)P'$, equality must hold in (1.4) and (1.5).

This situation is roughly akin to projecting onto subspaces in least squares theory. In particular, Csiszar defines an \mathcal{G} to be a linear set of PD's if $P, P' \in \mathcal{G}$ implies $\alpha P + (1-\alpha)P' \in \mathcal{G}$ for every α for which it is a PD. If \mathcal{G} is a linear set, then the inequality sign in (1.4) and (1.5) may be replaced by an equality sign as long as \mathcal{X} is finite. Based upon this characterization, Csiszar

is able to prove that if \mathcal{X} is finite, $\mathcal{S} = \bigcap_1^t \mathcal{S}_i$ is a finite intersection of arbitrary linear sets, and there exists a $P \in \mathcal{S}$ such that $I(P||R) < \infty$, then successive, cyclic, iterated I-projections onto the individual sets must converge to the I-projection of R onto \mathcal{S} . Thus if $Q_0 = R$, and Q_n denotes the I-projection of Q_{n-1} onto \mathcal{S}_n (where $\mathcal{S}_n = \mathcal{S}_i$ if $n = lt + i$, $1 \leq i \leq t$), then Q_n must converge to $Q \in \mathcal{S}$ as $n \rightarrow \infty$ where $I(Q||R) = \min_{P \in \mathcal{S}} I(P||R)$.

This result is very much dependent upon the assumption that the \mathcal{S}_i be linear sets (in fact it is not true in general) and the accompanying fact that equality holds in (1.4).

2. The Procedure

We now propose a procedure which will enable one to obtain I-projections onto a finite intersection of arbitrary closed, convex sets of PD's by iteratively finding I-projections onto the individual sets. We will prove that that under a mild restriction, the procedure must give the correct solution, and then examine an example.

First let us note that we can still define I-projections for non-negative vectors which are not normed to be PMF's onto sets of

PD's. That is, for $r \geq 0$ ($r \neq 0$), $Q \in \mathcal{G}$ is an I-projection of r onto a set of PD's \mathcal{G} iff

$$\sum_1^m q(k) \ln \frac{q(k)}{r(k)} = I(Q||R) = \min_{P \in \mathcal{G}} I(P||R) = \min_{P \in \mathcal{G}} \sum_1^m p(k) \ln \frac{p(k)}{r(k)}$$

and $I(Q||R) < \infty$.

Of course, since \mathcal{G} contains only PD's, the I-projection of r onto a convex set \mathcal{G} is identical to the I-projection of $r' = r / \sum_1^m r(k)$ (r normalized to be a PMF) onto \mathcal{G} . It is easily shown that the characterization of I-projections given in (1.4) and (1.5) is still valid, even though $I(P||R) \geq 0$ need no longer be true. (We note that multiplication and division of vectors will refer to the operations being performed coordinatewise.)

Let us now state our algorithm. We assume we wish to find the I-projection of r onto $\mathcal{G} = \bigcap_1^t \mathcal{G}_i$, where \mathcal{G}_i are closed, convex sets of PD's. We assume that we can project onto each \mathcal{G}_i individually and shall denote the I-projection of S onto \mathcal{G}_i by $\phi(\mathcal{G}_i|S)$. We also assume there exists a PD $T \in \mathcal{G}$ such that $I(T||R) < \infty$.

1. Let $S_{1,1} = R$, and let $\phi_{1,1} = \phi(\mathcal{G}_1|S_{1,1})$. We then set

$$s_{1,2} = p_{1,1} = r \frac{p_{1,1}}{s_{1,1}}. \quad (\text{We note that if } s_{1,1}(k) = 0, \text{ then}$$

so is $p_{1,1}(k)$. We take $0/0$ to be 1.)

2. Let $p_{1,2} = \phi(\mathcal{G}_2|S_{1,2})$. Set $s_{1,3} = p_{1,2} = r \frac{p_{1,1}}{s_{1,1}} \frac{p_{1,2}}{s_{1,2}}$.

3. Continue, until $p_{1,t} = \phi(\mathcal{G}_t|S_{1,t})$ where $s_{1,t} = p_{1,t-1} = r \frac{p_{1,1}}{s_{1,1}} \dots \frac{p_{1,t-1}}{s_{1,t-1}}$. Set $s_{2,1} = r \frac{p_{1,2}}{s_{1,2}} \dots \frac{p_{1,t}}{s_{1,t}}$.

Note that $s_{2,1} = p_{1,t} / \left(\frac{p_{1,1}}{s_{1,1}} \right)$.

$$4. \quad p_{2,1} = \varphi(\delta_1 | S_{2,1}). \quad \text{Set } s_{2,2} = r \frac{p_{2,1}}{s_{2,1}} \frac{p_{1,3}}{s_{1,3}} \dots \frac{p_{1,t}}{s_{1,t}},$$

or equivalently, $s_{2,2} = p_{2,1} / \left(\frac{p_{1,2}}{s_{1,2}} \right)$.

$$5. \quad \text{Continue. In general, set } s_{n,i} = p_{n,i-1} / \left(\frac{p_{n-1,i}}{s_{n-1,i}} \right), \quad 2 \leq i \leq t$$

and set $s_{n,1} = p_{n-1,t} / \left(\frac{p_{n-1,1}}{s_{n-1,1}} \right)$. We then let $p_{n,i} = \varphi(\delta_i | S_{n,i})$,

and define $S_{n,i+1} (S_{n+1,1} \text{ if } i = t)$, etc.

Suppose now that the δ_i are actually linear sets so that equality holds in (1.5). Noting that for any $p \in \delta_i$

$$(2.1) \quad I(p | S_{n,i}) = \sum_k p(k) \ln [p(k) / s_{n,i}(k)]$$

$$= \sum_k p(k) \ln [p(k) / (p_{n,i-1}(k) / p_{n-1,i}(k) / s_{n-1,i}(k))]]$$

$$(2.2) = \sum_k p(k) \ln \frac{p(k)}{p_{n,i-1}(k)} + \sum_k p(k) \ln \frac{p_{n-1,i}(k)}{s_{n-1,i}(k)},$$

we observe that the last term must be equal to

$$\sum_k p_{n-1,i}(k) \ln \left(\frac{p_{n-1,i}(k)}{s_{n-1,i}(k)} \right)$$

and hence free of p . Thus the $p \in \delta_i$ which minimizes (2.1), is also the one which minimizes the first part of (2.2), i.e., the I-projection of $p_{n,i-1}$ onto δ_i . It easily follows that our procedure reduces to the cyclic, iterative procedure given by Csiszar when the δ_i are closed linear sets.

For a simple example to show that Csiszar's procedure does not work for general convex sets, consider the following:

$$\mathcal{G}_1 = \left\{ \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}; p_{11} \geq p_{12}, p_{21} \geq p_{22}, p_{ij} \geq 0, \sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1 \right\},$$

$$\mathcal{G}_2 = \left\{ \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}; p_{11} \geq p_{21}, p_{12} \geq p_{22}, p_{ij} \geq 0, \sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1 \right\},$$

and

$$R = \begin{pmatrix} \frac{1}{16} & \frac{3}{16} \\ \frac{7}{16} & \frac{5}{16} \end{pmatrix}.$$

Csiszar's procedure yields $\begin{pmatrix} \frac{9}{32} & \frac{7}{32} \\ \frac{9}{32} & \frac{7}{32} \end{pmatrix}$, whereas $\begin{pmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix}$ is the correct solution.

Everything hinges on the following theorem.

Theorem 2.1. Assume $\mathcal{G} = \bigcap_1^t \mathcal{G}_i$, where the \mathcal{G}_i are closed, convex sets of PD's and $R \neq 0$ is a nonnegative vector such that there exists a $T \in \mathcal{G}$ where $I(T||R) < \infty$. If there exists a convergent subsequence $p_{n_j, i} \rightarrow p$ for some i such that

$$(2.3) \quad \liminf_j \sum_k (p_{n_j, i}^{(k)} - p(k)) \ln \left(\frac{p_{n_j, i}^{(k)}}{s_{n_j, i}^{(k)}} \right) \geq 0$$

for every i , then $p_{n, i} \rightarrow p$ as $n \rightarrow \infty$ and $P = \varphi(\mathcal{G}|R)$ where $p_{n, i}$ and $s_{n, i}$ are defined as in the proposed algorithm.

Proof. Recall that $p_{n, i} = \varphi(\mathcal{G}_i | s_{n, i})$ where

$$s_{n,i} = r \frac{p_{n,1}}{s_{n,1}} \dots \frac{p_{n,i-1}}{s_{n,i-1}} \frac{p_{n-1,i+1}}{s_{n-1,i+1}} \dots \frac{p_{n-1,t}}{s_{n-1,t}}.$$

$$(2.4) \quad \begin{cases} p_{n-1,t} \left(\frac{p_{n-1,1}}{s_{n-1,1}} \right) & \text{if } i = 1 \\ p_{n,i-1} \left(\frac{p_{n-1,i}}{s_{n-1,i}} \right) & \text{if } 2 \leq i \leq t. \end{cases}$$

Thus

$$\begin{aligned} & I(p_{n,i} \| s_{n,i}) - I(p_{n-1,i} \| s_{n-1,i}) \\ &= \sum p_{n,i} \ln \left(\frac{p_{n,i}}{s_{n,i}} \right) - \sum p_{n-1,i} \ln \left(\frac{p_{n-1,i}}{s_{n-1,i}} \right) \\ (2.5) \quad &= \sum p_{n,i} \ln \left(\frac{p_{n,i}}{p_{n,i-1}} \right) + \sum p_{n,i} \ln \left(\frac{p_{n-1,i}}{s_{n-1,i}} \right) - \sum p_{n-1,i} \ln \left(\frac{p_{n-1,i}}{s_{n-1,i}} \right) \\ &= \sum p_{n,i} \ln \left(\frac{p_{n,i}}{p_{n,i-1}} \right) + \sum (p_{n,i} - p_{n-1,i}) \ln \left(\frac{p_{n-1,i}}{s_{n-1,i}} \right) \\ &\geq I(p_{n,i} \| p_{n,i-1}) \quad \text{if } 2 \leq i \leq t, \text{ and similarly if } i = 1, \end{aligned}$$

since (by 1.4) the last term must be nonnegative because $P_{n,i} \in \mathcal{G}_i$.
 Noting that $I(P_{n,i} \| P_{n,i-1}) \geq 0$ since $P_{n,i}$ and $P_{n,i-1}$ are PD's,
 we have that $I(P_{n,i} \| S_{n,i})$ is nondecreasing in n for each i . Let
 us now show that these sequences are bounded above.

For $V \in \bigcap_1^t \mathcal{G}_1$,

$$\begin{aligned}
 I(V \| P_{n,i}) &= \sum v \ln \left(\frac{v}{p_{n,i}} \right) \\
 &= \sum v \ln v - \sum v \ln r \left[\frac{p_{n,1}}{s_{n,1}} \dots \frac{p_{n,i}}{s_{n,i}} \frac{p_{n-1,i+1}}{s_{n-1,i+1}} \dots \frac{p_{n,t}}{s_{n,t}} \right] \\
 (2.6) \quad &= \sum v \ln v - \sum v \ln r - \sum_{j=1}^t \left[\sum v \ln \frac{p_{a,j}}{s_{a,j}} - I(P_{a,j} \| S_{a,j}) \right] \\
 &\quad - \sum_{j=1}^t I(P_{a,j} \| S_{a,j}) \text{ where } a = \begin{cases} n, & j \leq i \\ n-1, & j > i. \end{cases} \\
 &\leq I(V \| R) - \sum_{j=1}^t I(P_{a,j} \| S_{a,j})
 \end{aligned}$$

by (1.4) and the fact that V belongs to every \mathcal{G}_i . Thus, choosing
 V such that $I(V \| R) < \infty$, we have a uniform upper bound on $I(P_{n,i} \| S_{n,i})$.

Hence the $\lim_{n \rightarrow \infty} I(P_{n,i} \| S_{n,i})$ exists finite for every i , and by (2.5)

$$(2.7) \quad I(P_{n,i} \| P_{n,i-1}) \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad \text{for } 2 \leq i \leq t.$$

(Similarly $I(P_{n,1} \| P_{n-1,t}) \rightarrow 0$ as $n \rightarrow \infty$.)

Note that (1.2) and (2.7) imply that if $p_{n_j,i} \rightarrow p$ for some i , then $p_{n_j,i} \rightarrow p$ for every i . Thus $p \in \bigcap_1^t \mathcal{G}_i$ since $p_{n_j,i} \in \mathcal{G}_i$ and the \mathcal{G}_i are closed.

Using (1.5) and (2.3), for any $v \in \bigcap_1^t \mathcal{G}_i$,

$$\begin{aligned} 0 &\leq \sum_i \liminf_j \sum_k (p_{n_j,i} - p) \ln \left(\frac{p_{n_j,i}}{s_{n_j,i}} \right) \\ &\quad + \sum_i \sum_k (v - p_{n_j,i}) \ln \left(\frac{p_{n_j,i}}{s_{n_j,i}} \right) \\ &\leq \liminf_j \sum_i \sum_k (v - p) \ln \left(\frac{p_{n_j,i}}{s_{n_j,i}} \right) \\ &= \liminf_j \sum_k (v - p) \ln r \prod_i \left(\frac{p_{n_j,i}}{s_{n_j,i}} \right) / r \\ &= \lim_j \sum_k (v - p) \ln \left(\frac{p_{n_j,t}}{r} \right) \\ &= \sum_k (v - p) \ln \left(\frac{p}{r} \right). \end{aligned}$$

Thus $P = \mathcal{G}(\mathcal{G}|R)$ by (1.5). Setting $V = P$ in (2.6) and using (2.3), it follows that for the subsequence $\{n_j\}$,

$$\sum_{i=1}^t I(P_{n_j, i} \| S_{n_j, i}) \rightarrow I(P \| R).$$

However, by the monotonicity in n of $I(P_{n, i} \| S_{n, i})$, we have that

$$\sum_{i=1}^t I(P_{n, i} \| S_{n, i}) \rightarrow I(P \| R).$$

Thus $I(P \| P_{n, t}) \rightarrow 0$, so that by (1.2) and (2.5), $p_{n, i} \rightarrow p$ as $n \rightarrow \infty$ for all i .

We remark that it would take rather surprising behavior of the $p_{n, i}$ for condition (2.3) to not hold, and we strongly conjecture that this condition is always true. As we note in the following corollary, if the $p_{n, i} \not\rightarrow P(\mathcal{S} | R)$, we must have $\sup_{n, i} \sum_k s_{n, i}(k) \rightarrow \infty$, and we have been unable to construct examples where this happens. We point out that when one uses the algorithm, they can put in a step to check the value of $\sup_{n, i} \sum_k s_{n, i}(k)$. If the algorithm is not going to converge correctly, then this value must become excessively large. Otherwise, the algorithm must converge to the correct solution.

Corollary 2.1. If the algorithm should not converge correctly, then

$$\sup_{n, i} \sum_k s_{n, i}(k) \rightarrow \infty \text{ as } n \rightarrow \infty.$$

Proof. Suppose $s_{n, i}(k)$ is uniformly bounded above, and $p_{n_j, i} \rightarrow p$ is a convergent subsequence (which must exist since $0 \leq p_{n, i}(k) \leq 1$). If there exists $0 < m \leq s_{n_j, i}(k)$ for all n_j, k , then (2.3) follows by standard continuity properties. If not, there exists at least one k such that $s_{n_j, i}(k)$ takes on arbitrarily small values. Then either

$p_{n_j,i}(k) \rightarrow 0$ for all such k (which cause no problems), or there exists a k such that $p_{n_j,i}(k) \rightarrow p(k) > 0$, while $s_{n_j,i}(k)$ takes on arbitrarily small values. This contradicts the uniform upper bound of $I(P_{n,i} \| S_{n,i})$.

We recommend that when one uses the algorithm, they should compute the average value over an entire cycle $(\frac{1}{t} \sum_{i=1}^t p_{n,i})$ rather than a single projection $p_{n,i}$ to estimate the I-projection. Convergence is still guaranteed, and this value is much more stable and seems to converge much more quickly to the correct solution.

3. An Example

To illustrate our algorithm, we consider $n \times n$ arrays of probabilities, say (p_{kj}) , $p_{kj} \geq 0$, $k, j = 1, \dots, n$, $\sum_{j=1}^n \sum_{k=1}^n p_{kj} = 1$. We denote the corresponding marginal PD's by $p_{k.} = \sum_{j=1}^n p_{kj}$ and $p_{.j} = \sum_{k=1}^n p_{kj}$, $k, j = 1, \dots, n$.

We now consider the problem of finding the I-projection of a fixed array (r_{kj}) subject to the marginal PD's being stochastically ordered, i.e.

$$\sum_{\ell=1}^i p_{\ell.} \geq \sum_{\ell=1}^i p_{. \ell}, \text{ for all } i.$$

(Kullback (1971) has given an iterative procedure for I-projections where equality is forced to hold for all i , also known as marginal homogeneity.

Equivalently, we want to find the I-projection of (r_{kj}) onto

$$(3.1) \quad \mathcal{G} = \left\{ \sum_{i=1}^{n-1} \phi_i \text{ where } \phi_i = \{(p_{kj}); \sum_{\ell=1}^i \sum_{m=1}^n p_{\ell,m} \geq \sum_{m=1}^i \sum_{\ell=1}^n p_{\ell,m}\} \right\}.$$

Note that the ϕ_i are closed, convex sets of PD's which are not linear sets. I-projections of $(v_{k,j})$ onto the ϕ_i can be found by forcing equality in the constraint if the $(v_{k,j})$ violate the constraint. (See Theorem 2.11 of Barlow et al. (1972) which can be modified to apply to arbitrary, convex functions.)

To express $P(\phi_i|V)$, we let

$$A_i = \{(\ell, m); 1 \leq \ell \leq i, i+1 \leq m \leq n\},$$

$$B_i = \{(\ell, m); 1 \leq m \leq i, i+1 \leq \ell \leq n\},$$

$$C_i = \{(\ell, m); \ell, m = 1, \dots, i\} \cup \{(\ell, m); \ell, m = i+1, \dots, n\}$$

$$\text{and } \hat{v}_i = \left(\sum_{A_i} v_{\ell, m} \sum_{B_i} v_{\ell, m} \right)^{1/2}.$$

If $(v_{k,j})$ satisfies the constraint of ϕ_i ($\sum_{A_i} v_{\ell, m} \geq \sum_{B_i} v_{\ell, m}$), then

$$P(\phi_i|V)_{k,j} = v_{k,j} / \sum_{\ell=1}^n \sum_{m=1}^n v_{\ell, m} \quad \text{for all } k, j.$$

If $(v_{k,j})$ does not satisfy the constraint of ϕ_i ($\sum_{A_i} v_{\ell, m} < \sum_{B_i} v_{\ell, m}$) then

$$\phi(\phi_i|V)_{k,j} = \begin{cases} v_{k,j} \left(\frac{\sum_{B_i} v_{\ell, m}}{\sum_{A_i} v_{\ell, m}} \right)^{1/2} [2 \hat{v}_i + \sum_{C_i} v_{\ell, m}]^{-1}, & (k, j) \in A_i, \\ v_{k,j} \left(\frac{\sum_{A_i} v_{\ell, m}}{\sum_{B_i} v_{\ell, m}} \right)^{1/2} [2 \hat{v}_i + \sum_{C_i} v_{\ell, m}]^{-1}, & (k, j) \in B_i, \\ v_{k,j} [2 \hat{v}_i + \sum_{C_i} v_{\ell, m}]^{-1}, & (k, j) \in C_i. \end{cases}$$

The key point is that finding the I-projection onto ϕ_1 is quite easy (and easily programmed), while finding the I-projection onto $\phi = \bigcap_{i=1}^{n-1} \phi_i$ is very difficult. However our algorithm enables one to find the latter I-projection using only the ability to handle the I-projections onto the individual ϕ_i .

To illustrate our example with some numbers, we consider some rather famous data from Stuart (1953) concerning grades of unaided distance vision for left and right eyes. If one wished to estimate the probabilities of falling into the various categories, subject to the provision that right eye vision is at least as good as left eye vision, one might find the I-projection of the data in Table 1 onto the ϕ given in (3.1). Using this algorithm, we have essentially obtained convergence to the true I-projection by 3 cycles. These values are listed in Table 2 (with the unrestricted MLE's given in parentheses).

This estimate might prove useful for constructing a likelihood ratio type test for testing whether right eye vision is better than left eye vision. This data is treated by Plackett (1981) who uses it for tests involving marginal homogeneity and quasi-symmetry. It also appears in Kendall (1974).

4. Acknowledgement

The author would like to thank a referee and the associate editor for helpful, knowledgeable comments and Peter Wollan for doing the calculations in Table 2.

Table 1
Unaided Distance Vision
 (From Kendall (1974)).

		LEFT EYE				
	Highest grade	Second grade	Third grade	Lowest grade	Totals	
RIGHT EYE	Highest grade	821	112	85	35	1053
	Second grade	116	494	145	27	782
	Third grade	72	151	583	87	893
	Lowest grade	43	34	106	331	514
	Totals	1052	791	919	480	3242

Table 2
I-Projection of Data in Table 1
 (Values in parentheses are Table 1 values normed
 to sum to unity.)

		LEFT EYE				
	Highest Grade	Second Grade	Third Grade	Lowest Grade	Totals	
RIGHT EYE	Highest grade	.2534 (.2532)	.0344 (.0345)	.0262 (.0262)	.0120 (.0108)	.3260 (.3247)
	Second grade	.0358 (.0358)	.1525 (.1524)	.0447 (.0447)	.0092 (.0083)	.2422 (.2412)
	Third grade	.0222 (.0222)	.0466 (.0466)	.1799 (.1798)	.0298 (.0268)	.2785 (.2754)
	Lowest grade	.0120 (.0133)	.0095 (.0105)	.0295 (.0327)	.1022 (.1021)	.1532 (.1586)
	Totals	.3234 (.3245)	.2430 (.2440)	.2803 (.2834)	.1532 (.1480)	.9999 (.9999)

REFERENCES

1. Barlow, R.E., Bartholomew, D.J., Bremner, J.M., and Brunk, H.D. (1972). Statistical Inference Under Order Restrictions. Wiley, New York.
2. Cencov, N.N. (1972). Statistical Decision Rules and Optimal Inference. Translation of Mathematical Monographs, American Mathematical Society, Vol. 53.
3. Csiszar, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* 2 299-318.
4. Csiszar, I. (1975). I-divergence geometry of probability distributions and minimization problems. *Ann. Prob.* 3 146-159.
5. Good, I.J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* 34 911-934.
6. Jaynes, E.T. (1957). Information Theory and Statistical Mechanics. *Phys. Rev.* 106 620-630.
7. Jeffreys, H. (1948). Theory of Probability. 2nd ed. Clarendon Press, Oxford.
8. Kemperman, J.H.B. (1967). On the optimum rate of transmitting information. Probability and Information Theory. Lecture Notes in Mathematics, Springer-Verlag, 126-169.
9. Kendall, M.G. (1974). Rank Correlation Methods. Charles Griffin and Co., London.
10. Kullback, s. (1959). Information Theory and Statistics. Wiley, New York.

11. Kullback, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory*, IT-13 126-127.
12. Kullback, S. (1971). Marginal homogeneity of multidimensional contingency tables. *Ann. Math. Statist.* 42 594-606.
13. Meyer, Michael M. (1980). Generalizing the iterative proportional fitting procedure. Dept. of Applied Statistics Tech. Report No. 371, Univ. of Minnesota.
14. Plackett, R.L. (1981). The Analysis of Categorical Data. Macmillan,
15. Rao, C.R. (1965). Linear Statistical Inference and its Applications. Wiley, New York.
16. Sanov, I.N. (1957. On the probability of large deviations of random variables. *Mat. Sbornik* 42 11-44.
17. Stuart, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika*, 40 105-110.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 106	2. GOVT ACCESSION NO. AD-A145402	3. RECIPIENT'S CATALOG NUMBER 6
4. TITLE (and Subtitle) An Iterative Procedure for Obtaining I-projections onto the Intersection of Convex Sets.		5. TYPE OF REPORT & PERIOD COVERED August - 1984
6. AUTHOR(s) Richard L. Dykstra		7. PERFORMING ORG. REPORT NUMBER 106
8. CONTRACT OR GRANT NUMBER(s) N00014-83-K-0249		9. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
10. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics and Actuarial Science The University of Iowa Iowa City, Iowa 52242		11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy Arlington, VA 22217 -
12. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. DATE August 1984
		14. NUMBER OF PAGES 19
		15. SECURITY CLASS. (of this report) Unclassified
		16. DECLASSIFICATION/DOWNGRADING SCHEDULE
17. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.		
18. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
19. SUPPLEMENTARY NOTES		
20. KEY WORDS (Continue on reverse side if necessary and identify by block number) I-divergence, I-projections, convexity, Kullback-Liebler information number, cross entropy, iterative projections, iterative proportional fitting procedure.		
21. ABSTRACT (Continue on reverse side if necessary and identify by block number) A frequently occurring problem is to find that probability distribution (PD) lying within a set \mathcal{C} which minimizes the I-divergence between it and a given PD R . This is referred to as the I-projection of R onto \mathcal{C} . Csizsar (1975) has shown that when $\mathcal{C} = \bigcap_{i=1}^t \mathcal{C}_i$ is a finite intersection of closed, linear sets, a cyclic, iterative procedure which projects onto the individual \mathcal{C}_i must converge to the desired I-projection on \mathcal{C} providing the sample space is finite. (cont.)		

END

FILMED

10-84

DTIC